DOCUMENT RESUME

ED 369 226 EC 302 956

AUTHOR	Thurlow, Martha L.; Ysseldyke, James E.
TITLE	Can "All" Ever Really Mean "All" in Defining and Assessing Student Outcomes? Synthesis Report 5.
INSTITUTION	National Association of State Directors of Special Education, Alexandria, VA.; National Center on Educational Outcomes, Minneapolis, MN.; Saint Cloud State Univ., MN.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	Mar 93
CONTRACT	H159C00004
NOTE	18p.; Paper presented at the Annual Meeting of the American Evaluation Association (Seattle, WA, November 5, 1992).
PUB TYPE	Speeches/Conference Papers (150) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Standards; Definitions; *Disabilities; *Educational Policy; Elementary Secondary Education; *Equal Education; Evaluation Methods; *Outcomes of Education; *Student Evaluation

ABSTRACT

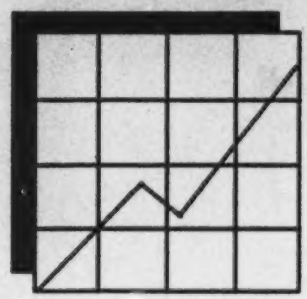
This paper cites evidence that the word "all" in referring to the education of all children is not being used in the universal sense, and notes that actions have not matched the rhetoric about "all" students. The paper defines the terms "outcomes," "indicators," and "standards." It discusses key issues that arise in talking about outcomes for all students--practical issues, technical issues, legal issues, and philosophical issues. It describes ways in which "all" really can mean "all" in defining student outcomes, citing the examples of Kentucky and Arizona which include all students in evaluation of educational outcomes. The paper notes that most disability groups want the same accommodations allowed during testing as society allows for people with disabilities to live in the community and to work, while educators have not yet reached this level of consensus on the issue of whether the education of students with disabilities should strive for the same outcomes as the education of students without disabilities. The paper concludes with two qualifiers for "all means all" : outcomes must be relevant to all, and there is a need to identify innovative ways to assess universal outcomes. (Contains 28 references.) (JDD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 369 226

Synthesis Report 5



Can "All" Ever Really Mean "All" in Defining and Assessing Student Outcomes?

National Center on Educational Outcomes

The College of Education
UNIVERSITY OF MINNESOTA

in collaboration with

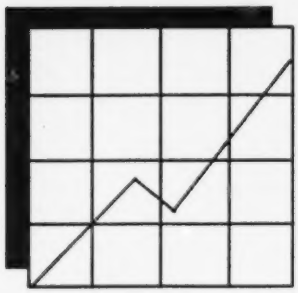
St. Cloud State University
and
National Association of State Directors of Special Education

March, 1993

BEST COPY AVAILABLE

EC 302956

Synthesis Report 5



Can "All" Ever Really Mean "All" in Defining and Assessing Student Outcomes?

Martha L. Thurlow, and James E. Ysseldyke

National Center on Educational Outcomes

This report is a paper that was presented on November 5, 1992, at the annual meeting of the American Evaluation Association in Seattle, Washington. The development of this paper was supported through a Cooperative Agreement (H159C00004) with the U.S. Department of Education, Office of Special Education Programs. Opinions or points of view do not necessary represent those of the U.S. Department of Education or Offices within it.

March, 1993

Can "All" Ever Really Mean "All" in Defining and Assessing Student Outcomes?

Martha L. Thurlow and James E. Ysseldyke
National Center on Educational Outcomes

Paper Prepared for the Annual Meeting of the American Evaluation Association

We have been asked to answer the question, "Can 'all' every really mean 'all' when defining student outcomes?" Immediately, our inclination is to respond: "Yes . . . and no," or "It depends on your perspective." But, we are not satisfied with these responses. We must recognize that there is no easy answer, yet at the same time not use that as a cop out to fail to take a stand.. We also must realize that any stance we take will have drawbacks and limitations. We will argue that "all means all," but that accepting this answer will require some adjustments in traditional thinking. In order to develop our response more fully, it is necessary to explore several topics.

Why Do We Even Ask This Question?

To a great extent, we ask the question "Can 'all' ever really mean 'all'?" because of recent educational and political events in our country. Within the past few years, we have repeatedly heard the phrases "all children," "all students," and "all American citizens." What does "all" mean? According to Webster, all is "the whole amount or quantity"; it is "every member or individual component of" some group or thing. "All" is "every." It is "the whole number or sum of." "All" is the universal quantifier.

"All" is a term used in much rhetoric about education in our nation today -- all students have the right to a free and appropriate public education, all children must meet world class standards, all citizens must be literate and lifelong learners, and so on. But, several important events and public documents in the past few years have suggested that "all" is being given a meaning different from the dictionary meanings presented by Webster.

For example, when proposing the agenda for education at the Education Summit Meeting in Charlottesville Virginia, President Bush said that in striving to reach the nation's goals for education, there was a need to ensure that "no child in America be forgotten or forsaken" and that this "includes both the unusually gifted and those with special needs and disabilities" (White House, 1990). But, did all really mean all to President Bush and the governors who helped formulate the goals? It doesn't seem so. The six national education goals that were identified stressed academic achievement and academic excellence. Then, when identifying sources of data for monitoring progress on the national goals, the National Education Goals Panel (1991) identified data collection programs that, for the most part, exclude from 40-50% of students with disabilities (McGrew, Thurlow, Shriner, & Spiegel, 1992).

Before this, the National Council of Teachers of Mathematics (1987) proposed standards for mathematics instruction, emphasizing that the standards applied to all students. Yet, there was no attempt to consider students with disabilities when the standards were developed (Shriner, Kim, Thurlow, & Ysseldyke, 1992a). And, there are mixed perceptions of how useful the standards are for the education of many students with disabilities (Shriner, Kim, Thurlow, & Ysseldyke, 1992b).

In 1992, the National Council on Educational Standards and Testing (NCEST) produced Raising Standards for American Education. The Council had been established to address issues surrounding Goals 3 and 4, the goals on student academic achievement and being first in the world in science and mathematics. In its report, NCEST endorsed the concept of national standards for all students. The Council also stated:

> The Council's intent in recommending the establishment of national standards is to raise the ceiling for students who are currently above average and to lift the floor for those who now experience the least success in school, including those with special needs. (p. 4)

Although applauded in general, the Raising Standards document generated considerable debate. The Council for Exceptional Children (1992) specifically noted the following issues in testimony before the House Subcommittee on Elementary, Secondary and Vocational Education:

- Even though standards are considered to be "world class," they must accommodate all students. "We cannot accept . . . students with severe disabilities [being] 'exempt' from meeting world class standards."
- If standards are set only in English, mathematics, history, science, and geography, this might lead to a narrowing of the curriculum, which would have a negative impact on students with disabilities.
- If national assessments are developed, they must include all students. There is a need to investigate alternative forms of assessment as well as ways to modify existing assessments.
- If a group is formed to certify standards and criteria for assessment, it must include at least one member who has expertise in working with individuals with disabilities.

These issues suggested that the term all was not being used by NCEST in a way that really meant all students being educated within the schools of today.

There is other evidence that the word "all" is not being used in the universal sense in much of the educational rhetoric of today. A major reform approach in the 1990s has been the movement toward outcomes-based education. Spady and his associates in the High Success Network have been among the major proponents of a new vision of success for all learners, with instruction based on desired outcomes rather than restricted by existing curricula, time constraints, or other traditional barriers in education (Spady, 1992). Typical exit outcomes delineated by schools adopting Spady's vision include such ideals as being a competent, productive participant in society, making decisions for successful living, applying knowledge in diverse situations, and learning over a lifetime. Instruction is not bound by the existing curriculum or schedules, but rather is designed to meet the individual needs of each student. The implementation of these ideals, however, has been difficult for numerous local school districts. We hear too often that students with disabilities are being excluded as outcomes are identified and instructional programs developed.

Actions have not matched the rhetoric about "all" students. The term "all" is being used by many groups because it sounds good to be inclusive of all students.

However, when the time comes to operationalize proposals and initiatives, few of the "all" proponents are dealing with the implications of the inclusive language. Measurement groups repeatedly develop lists of reasons why it is very difficult to accommodate students with disabilities in state and national testing programs (for example, see the draft paper of the APA Division 15 Task Force, 1992). Advocates for people with disabilities are having to fight for the attention of groups setting the national education agenda. Repeated evidence of the disparity between talk of including all students and actual practice led Stevan Kukic, as President of the National Association of State Directors of Special Education (NASDSE), to state emphatically "ALL MEANS ALL."

The question we are addressing -- "Can all ever really mean all?" -- remains a prominent question in the minds of many when it comes to student outcomes. We will address this question by clarifying some key terms, identifying and discussing several important issues, and then delineating ways in which all can mean all when it comes to defining and assessing student outcomes.

Definitions of Terms

Several terms related to outcomes, including the term "outcomes" itself have been used to mean different things by different people. At the same time, different terms sometimes are used to mean the same things. Three terms that seem to cause the most confusion are "outcomes," "indicators," and "standards."

Let's start with standards. This term, though used extensively in Raising Standards for American Education, was not defined in that document other than through examples. Many of the examples were from activities that individual states had completed, and exemplified the NCEST notion that student standards include "specification of the content -- what students should know and be able to do -- and the level of performance that students are expected to attain -- how good is good enough" (p. 3). Among the examples provided were Maine's Common Core of Learning:

> Students with a common core of knowledge . . .
> - Are familiar with contemporary and enduring works of American literature and have a sense of how important themes of American experience have developed through time
> - Are familiar with works of diverse literary traditions -- works by women and men of many racial, ethnic, and cultural groups in different times and parts of the world, including Shakespeare, the Bible as literature, and classical mythology
> - Communicate clearly -- orally, in writing, and with graphics
>   -- Have a strong command of standard oral and written language conventions
>   -- Demonstrate basic proofreading and editing skills
>   -- Use handbooks and reference books to locate language terminology and rules (NCEST, 1992, p. 22)

and California's History-Social Science Framework:

> We want our students to understand the value, the importance, and the fragility of democratic institutions . . . to develop a keen sense of ethics and citizenship, and to care deeply about the quality of life in their community, their nation, and their world. (NCEST, 1992, p. 25)

In a recent report from the Office of Technology Assessment (1992), Testing in American Schools: Asking the Right Questions, the term "standards" was defined as follows:

> The word *standards* applied to tests has at least two different meanings. In the more general context it denotes goals, desirable behaviors, or models to which students, teachers, or schools should aspire. Such standards describe what optimal performance looks like and what is desirable for students to know. For example, the National Council of Teachers of Mathematics has determined that a standard for mathematics instruction is to emphasize mathematics as problem solving. The word standards, in its more technical meaning, denotes the specific levels of *proficiency* that students are expected to attain. Thus, setting a passing score for a test is equivalent to setting a standard of performance on that test. (p. 3)

Both of these definitions are useful.

In the recent past, outcomes were defined as the "results of education," typically including knowledge, skills, and attitudes. The notion was that schooling was to produce youngsters and young adults with certain levels of knowledge, certain types of skills, and certain attitudes about learning, life, and other people. In a more academic vein, NCEO defined "outcome" formally as "the result of interactions between individuals and schooling experiences" (Ysseldyke, Thurlow, Bruininks, Deno, McGrew, & Shriner, 1991). Later in its development of a model of outcomes, NCEO assigned this definition to the term "educational outcome" and then generated a new term, "enabling outcome," which is defined as "the result of interactions between individuals and life experiences that provide the individual with the opportunity to reach educational outcomes" (Ysseldyke, Thurlow, Bruininks, Gilman, Deno, McGrew, & Shriner, 1992).

In the development of its model of educational and enabling outcomes, NCEO initially delineated the educational outcomes as:

- Literacy
- Self Dependence
- Social/Behavioral
- Contribution/Citizenship
- Satisfaction
- Physical Health

and the enabling outcomes as:

- Presence/Participation
- Compensatory and Accommodation Skills
- Adaptive Behavior
- Family Coping and Support Skills

In a refinement of this "model" of outcomes, it identified and defined six educational outcomes:

- **Literacy** -- the use of information to function in society, to achieve goals, and to develop knowledge
- **Independence/Responsibility** -- the extent to which the individual's behavior reflects the ability to function independently and assume responsibility for one's self
- **Social/Behavioral Skills** -- the extent to which the individual demonstrates socially acceptable behavior
- **Contribution/Citizenship** -- the ways in which an individual gives something back to society or participates as a citizen in society
- **Satisfaction** -- degree to which a favorable attitude is held toward education
- **Physical/Mental Health** -- the extent to which the individual demonstrates healthy behavior, attitudes, and knowledge toward both physical and mental well-being

and two enabling outcomes:

- **Presence/Participation** -- the extent to which an individual is present in a particular setting and the extent to which meaningful participation occurs
- **Accommodation/Adaptation/Compensation** -- modifications that must be made to adjust to or make up for some type of disability

These "outcomes" are really outcome domains. They do not include statements of outcomes. The NCEO currently is working on a revision of the model that will include specific outcomes statements as well as key indicators of the outcomes.

An indicator was defined by Shavelson, McDonnell, and Oakes (1989) as "an individual or composite statistic that relates to a basic construct in education that is useful in a policy context." With the assumption that an indicator might be a qualitative statement, NCEO defined an indicator as "a symbolic representation of one or more outcomes . . . that can be used in making comparisons" (Ysseldyke et al., 1991). Indicators are often the statistics derived from assessments of students with respect to a particular outcome.

The three terms, "outcomes," "indicators," and "standards," form the basis for our stance that "all" means "all" when it comes to identifying and assessing students outcomes. Before we pursue this further, we must look at the major issues that generate the current quandary about the extent to which "all" means "all."

Issues in Identifying Outcomes for All

There are a number of key issues that arise as we talk about outcomes for all students. We label these, for simplicity, practical, technical, legal, and philosophical issues. They all play a part in our discussion of the question "Can all ever really mean all in defining student outcomes?"

Practical issues. Given the current framework within which most educators (particularly, educational administrators) typically operate, it seems to be easier not to include students with disabilities when thinking about educational outcomes. This perception may have come about as a result of circumstances related to the history of special education. At the time when discussion of the questionable results of education started, back in 1983 with the publication of A Nation at Risk (National Commission on Excellence in Education, 1983), students with disabilities generally were receiving extensive "special" education services. Many were excluded entirely from general classrooms, receiving their instruction instead in special classrooms, often in another part of the building. Many were in separate buildings. As we have said before, "out of sight" frequently turns into "out of mind" (NCEO, 1992), and this probably was the case when our nation began to question the results of education. Why worry about students with disabilities? They had their own educational system, with its associated mechanism for identifying objectives of instruction and a schedule for re-evaluating progress. In many cases, special educators had developed a special curriculum for these students. It seemed more practical to focus on students in general classrooms for whom the same basic curricula were being used. The others were "different."

Technical issues. Traditionally, the next step after identifying outcomes is the identification of outcome indicators and ways to measure the outcomes. The measurement of outcomes with typical general education assessment tools frequently is difficult when students with disabilities are included. Generally, some type of accommodation or adaptation is needed to compensate for the student's disability. Common types of accommodation and adaptation include modifications in presentation or response format, or the use of flexible timing or flexible setting arrangements. We have little empirical evidence on the effects of these arrangements on the reliability and validity of the assessment tools. In fact, the only studies that are available are for the Scholastic Aptitude Test (SAT) and the Graduate Record Examinations (GRE) conducted by ETS in the mid 1980s (see Willingham, Ragosta, Bennett, Braun, Rock, & Powers, 1988). The major concern of the researchers in these studies was the extent to which the scores from modified tests or accommodated testing situations were comparable to the scores from non-changed assessments. Here are some of the conclusions:

- Nonstandard and standard tests of both the SAT and the GRE had equivalent reliability (Bennett, Rock, & Kaplan, 1988).

- When supplemented with grade point averages, nonstandard tests did not consistently over or under-predict academic performance. The predictability of the academic performance of different subgroups of students with disabilities did vary (Braun, Ragosta, & Kaplan, 1986a, 1986b).

- Nonstandard and standard versions of the SAT and GRE were not comparable in terms of time to finish the test, but the noncomparability varied as a function of disability category (e.g., students taking special administrations of the SAT were more likely to complete it than those taking standard administration, while students with visual disabilities taking timed national administrations of the GRE were less likely to complete) (Bennett, Rock, Kaplan, & Jirele, 1988).

The ETS studies, while providing some information on the technical issues, include only two tests, ones generally used to make predictions about students.

Studies of minimum competency tests, which generally are used to make eligibility decisions, have identified issues, but for the most part have not looked at the actual effects of using modified versions of a test. Although there is a paucity of research in this area, at least one study examined the effects of test modifications on the minimum competency performance of students with learning disabilities (Beattie, Grise, & Algozzine, 1983). In this research, the format of the test was modified in various ways (e.g., items measuring similar skills were placed in progressive order of difficulty; unjustified margins were used on the reading comprehension section so that sentences did not have to be broken). In addition, one version of the test was printed in 12-point type rather than 10-point, and one version was printed in "large print" (18-point type). Format modifications appeared to enhance the test performance of students with learning disabilities, whereas print size did not appear to make a significant difference in performance. The extent to which such modifications alter the constructs being measured is certainly an issue that will be raised when reviewing these results and others that systematically examine the effects of testing accommodations and modifications.

In January 1992, individuals responsible for state assessment programs gave us input on what they believe are the major technical issues involved in the assessment of the educational outcomes of students with disabilities. They identified two primary issues:

(1) Inclusion of students with disabilities in data collection efforts.
(2) The nature and extent of testing modifications that are permitted.

The first issue deals with the criteria that are used to include (or more likely, exclude) students from assessment programs. It goes beyond mere written criteria to issues surrounding how the criteria actually are implemented and not implemented. The second issue deals with allowable testing accommodations and adaptations. Both of these issues encompass a wide array of other secondary technical issues. The state assessment personnel identified an array of issues that we were able to organize into 12 different categories . These issues, and a brief description of each, are presented in Table 1.

[Insert Table 1 about here]

Legal issues. Federal laws form the basis for several additional issues related to the comparability of educational outcomes for individuals with and without disabilities. These were summarized by Phillips (1992) in her discussion of legal challenges to decisions to not allow testing accommodations for students with disabilities. Constitutional due process protections of the fourteenth amendment apply when the results of testing will deprive the person of property or other interest, such as a high school diploma. The Individuals with Disabilities Education Act (IDEA), formerly the Education for All Handicapped Children Act (EHA), generally is not seen as a basis for allowing testing accommodations, because, according to the case cited by Phillips, EHA was intended to open the door of public education rather than to guarantee a particular level of education. On the other hand, the Americans with Disabilities Act (ADA), which was formerly covered by Section 504 of the Rehabilitation Act of 1973, may be more demanding in requirements for accommodations for individuals with disabilities. The actual impact of ADA on schools is uncertain because there is as yet no case law to which to refer.

The premise of most of Phillips's (1992) discussion is not whether the content of tests should be altered, but rather what changes can be made in testing conditions to accommodate a student's disabilities. Phillips noted the finding of U.S. Supreme Court

Justice Powell in 1979: "Section 504 imposes no requirement upon an educational institution to lower or to effect substantial modifications of standards to accommodate a handicapped person" (Southeastern Community College v. Davis). This decision suggests that the same outcomes are expected of all students. Phillips reinforces this notion by suggesting the ADA will be similar to Section 504 in its interpretation. But, this remains to be seen.

Philosophical issues. Inclusion is a philosophical issue that is at the heart of our question "can all ever really mean all?" True inclusion incorporates all aspects of life. This means that students with disabilities are included when our nation defines its educational goals. In turn, this means that if a specific assessment system is used to assess progress on our national education goals, then that system must include students with disabilities. If some of those students have disabilities that preclude their participation with appropriate testing accommodations, then their score should be entered as zero. That is the philosophy behind the "all means all" statement.

From a policy perspective, there is another philosophical reason for including students with disabilities in the identification of outcomes and ways to measure them:

> It is imperative that outcomes for these students be considered in the development of the larger outcome assessment systems. Special education as a program cannot exist as a separate and parallel program only loosely coupled with the larger system. The larger system must be accountable for what happens to these students and not simply for providing services to them. (McLaughlin & Warren, 1992, p. 31)

This argument is similar to that made by Nicholas Hobbs (1975) more than 15 years ago, when he noted that social and education agendas for students with special needs must be joined. Bruininks, Thurlow, and Ysseldyke (1992) concluded that there is a need for the "outcomes we seek from our schools [to] measure truly important educational results for all the nation's children and youth" (p. 99). Patricia McGill-Smith, an advocate and parent, likened reform and the goals of America 2000 to a train:

> Parents of children with disabilities are on the right platform, but we need tickets to get on the train. Outcomes data that help us evaluate reforms based on what works for our children are the tickets we need. (Ysseldyke, Thurlow, & Shriner, p. 49)

Whether we want to admit it or not, the philosophical issues are a key part of the answer we reach when we are asked about the inclusiveness of the term "all."

Ways in Which All Means All in Defining Student Outcomes

There are several ways in which "all" really can mean "all" in defining student outcomes. The current recommendation being made is to say simply that every outcome is for every student. In assessment of outcomes, this means that we should give a score of zero to any student who did not actually participate in the assessment. This is the approach that was used in the 1992 National Adult Literacy Survey (Kolstad, personal communication, July 1992). It is an approach that is likely to decrease the enormous amount of exclusion of students with disabilities that now occurs in both state and national data collection programs (see McGrew et al., 1992).

Other approaches also have been recommended or used. Some of these are suggested by what states do, others by educators' opinions, and still others by what representatives of the disability community say.

Two state examples. Recently, the state of Kentucky had the advantage of completely overhauling its educational system as part of its educational reform efforts. It may seem strange to call the upheaval of its educational system an advantage, but it is. Kentucky had to completely rethink the organization of its educational system. In doing so, it embraced an inclusive educational philosophy that permeated its management structure. Kentucky began by identifying educational outcomes for all students in its educational system. It determined that all students likewise would be included in the evaluation of the extent to which educational outcomes were reached.

In Kentucky, the desired educational outcomes are that all students will be able to:

(1) Use basic communication and math skills for purposes and situations they encounter in life
(2) Apply core concepts and principles from mathematics, the sciences, arts and humanities, social studies, practical living studies and vocational studies for purposes and situations they encounter in life,
(3) Become self-sufficient individuals
(4) Become responsible members of a family, work group, or community
(5) Think and solve problems across the variety of situations they encounter in life
(6) Connect and integrate the knowledge they have gained in school into their own lives.

These learning goals are presented as goals for all students. Within the learning goals are valued outcomes -- 75 in all. These also are presented as outcomes for all students. For example, in writing, the one valued outcome is that "students communicate ideas and information to a variety of audiences for a variety of purposes in a variety of modes through writing." In science, one of the six valued outcomes is that "students use appropriate and relevant scientific skills to solve specific problems in real life situations." Another is that "students use models and scales to explain or predict the organization, function, and behavior of objects, materials, and living things in their environment."

Student performance on these learning goals and valued learner outcomes is defined in terms of four performance levels (novice, apprentice, proficient, and distinguished), which are intended to be more useful and more positive than using letter grades. The definition of a range of performance levels is the first aspect of Kentucky's plan to identify goals and outcomes for all students. The second aspect of the Kentucky plan is the recognition that the assessment system designed to monitor progress on the six learning goals and 75 valued outcomes needs to be modified for persons with significant cognitive impairments. The desired outcomes are not being altered, just the means by which they are assessed. The assessment system, which involves the use of portfolios, is being field tested at this time.

Arizona is another state in which the state assessment system is being transformed so that all students with IEPs are included in the program. As Koehler (1992) recently noted:

> The removal of special education students from the "accountability track" also resulted, to a large degree, in their removal from the "curriculum track," those learning expectations which guided the instruction of regular education students. . . . This results in the special education student becoming more and more isolated from the mainstream instructional program rather than having an alternate course being charted for reaching competence in the mainstream subject area content. While the blame for such practice cannot be attributed solely to statewide standardized testing programs and special education students' exclusion from them, there is a strong case to be made that the absence of anchoring assessments to statewide standards has contributed significantly to the problem. (pp. 2-3)

These points have been addressed by others as well when they argue that out-of-sight too often means out-of-mind.

What disability groups want. We have met with representatives of a variety of disability groups. While they do not all agree, most express the opinion that they do not want to be excluded from the national agenda simply because of their disabilities. They want the same accommodations allowed during testing as society allows for people with disabilities to live in the community and to work. They have told us that if students are to be excluded from national or state assessments simply because they have disabilities, then they want them to be given zero scores in the data base. As noted previously, this strategy is one that likely will get the attention of policymakers and citizens, as a whole new waterfront of low scores appears.

Of course, there are others representing the disability community who do not want students with disabilities to be held to the same standards as other students. They believe that students with disabilities have unique educational needs that require the identification of a separate set of outcomes. In some cases, the argument is for a slightly modified set of outcomes, a set of outcomes that is an add-on to the general outcomes of education that apply to all students. The argument here is that students with disabilities have to learn much more than other students.

What educators say. The National Center on Educational Outcomes commissioned papers about inclusion and exclusion criteria for participation in national and state assessments, and on the potential effects of various testing accommodations. These issues, by implication, encompass the larger issue of whether the education of students with disabilities should strive for the same outcomes as the education of students without disabilities. Even among the few experts that have responded thus far, there is little consensus. For example, one person with extensive experience in measurement and evaluation has noted that the exclusion of students with disabilities, because of their small number, will have little effect on aggregated national or state assessment results. This person identifies the issue of construct validity (is the same construct measured for different subgroups?) as the critical issue to address. Another person has argued that we need to first identify areas of learning that are essential for thriving as human beings, areas so essential that we need to know about the progress of everyone on them. Then, we need to assess all students in these areas, although not necessarily expect the same level of achievement from all students. Nor should we expect that a single assessment instrument could be used. This person recognizes that we are facing a significant technical challenge.

Conclusions

Although we have argued for "all" really meaning "all" in defining student outcomes, we do so with the recognition that there are additional points that need to be made. First, and foremost, is the need to look critically at the outcomes that are defined. Are we identifying essential outcomes? Are we defining outcomes that are broad enough to be relevant for all students? This is a critical question, because we do see various groups defining outcomes that are very narrow in scope. We need a statement of what the critical outcomes of education are. If we have additional outcomes that we want to define for certain groups of students, then we need to be clear that that is what we are doing. We should not be claiming that there are outcomes for all students. This does not mean that the indicators of the outcomes are necessarily the same, nor that the same standards are held for all students. But it does mean that we define our critical outcomes first, outcomes for all students. Thus, our first qualifier for "ALL MEANS ALL" is that our outcomes must be relevant to all.

Second, we have a long way to go in exploring what an assessment system should be. We certainly cannot be limited to a single standardized test to measure progress on outcomes. We must think broadly as we think about the techniques of assessment. It may be that we will need a set of criteria for determining what type of assessment a student will take. For example, most students may participate in a standardized test, plus a few performance items (such as NAEP is now using), while others may participate in an alternate form of assessment (which could span the range from performance items, to a portfolio assessment, to even a knowledgeable informant assessment). At this point, we should not limit ourselves to a single format. We also need to be exploring more carefully ways to obtain comparable measures from different forms of assessment. We need to explore ways to develop technically adequate forms of alternative assessment. Thus, our second qualifier for "ALL MEANS ALL" is that we need to identify innovative ways to assess our universal outcomes. This means lots of hard work for lots of people. But, we believe that it is work that needs to be done. The time is right to be doing this work.

The current words of educational reform (world class standards, national tests, break-the-mold schools) have an aura of being appropriate for a small portion of the student population. We are not talking just about students with disabilities when we address the issue of "all means all" in American education today. We really are talking about the 25-30% of students for whom higher standards, world class standards, and other avenues to excellence make little sense. Out nation in its quest to become first in the world has forgotten many of its students. It is too easy to exclude larger and larger numbers of students from our educational system if we don't begin from "ALL MEANS ALL."

References

APA Division 15 Task Force. (1992). Draft of a statement of the American Psychological Association Division of Evaluation, Measurement, and Statistics on psychometric and assessment issues raised by the Americans with Disabilities Act (ADA). Unpublished paper.

Beattie, S., Grise, P., & Algozzine, B. (1983). Effects of test modifications on the minimum competency performance of learning disabled students. Learning Disability Quarterly, 6 (1), 71-77.

Bennett, R. E., Rock, D. A., & Kaplan, B. A. (1988). Level reliability and speededness of SAT scores for nine handicapped groups. Special Services in the Schools, 4 (3/4), 37-54.

Bennett, R. E., Rock, D. A., Kaplan, B. A., & Jirele, T. (1988). Psychometric characteristics. In W. W. Willingham, M. Ragosta, R. E. Bennett, H. Braun, D. A. Rock, & D. E. Powers (Eds.), Testing handicapped people (pp. 83-97) Boston: Allyn & Bacon.

Braun, H., Ragosta, M., & Kaplan, B. (1986a). The predictive validity of the GRE general test for disabled students. (ETS Research Report 86-42). Princeton, NJ: Educational Testing Service.

Braun, H., Ragosta, M., & Kaplan, B. (1986b). The predictive validity of the Scholastic Aptitude Test for disabled students. (ETS Research Report 86-38). Princeton, NJ: College Entrance Examination Board, Educational Testing Service.

Bruininks, R., Thurlow, M. L., & Ysseldyke, J. E. (1992). Assessing the right outcomes: Prospects for improving education for youth with disabilities. Education and Training in Mental Retardation, 27 (2), 93-100.

Council for Exceptional Children. (1992). Statement prepared for testimony before the House Subcommittee on Elementary, Secondary, and Vocational Education. Reston, VA: Author.

Hobbs, N. (1975). The futures of children. San Francisco: Jossey-Bass.

Koehler, P. H. (1992). The assessment of special needs students in Arizona. Unpublished manuscript.

McGrew, K. S., Thurlow, M. L., Shriner, J. G., & Spiegel, A. N. (1992). Inclusion of students with disabilities in national and state data collection programs (Technical Report 2). Minneapolis: National Center on Educational Outcomes, University of Minnesota.

McLaughlin, M. J., & Warren, S. H. (1992). Outcomes assessment for students with disabilities: Will it be accountability or continued failure? Preventing School Failure, 36 (4), 29-33.

National Commission on Excellence in Education. (1983) A nation at risk: The imperative for educational reform. Washington, DC: U. S. Government Printing Office.

National Council of Teachers of Mathematics. (1989) Curriculum and evaluation standards for school mathematics. Reston, VA: Author

National Education Goals Panel. (1991). The national education goals report: Building a nation of learners. Washington, DC: Author.

NCEST. (1992). Raising standards for American education. Washington, DC: Author.

Office of Technology Assessment (1992), Testing in American schools: Asking the right questions. Washington, DC: Author.

Phillips, S. E. (1992). Testing condition accommodations for handicapped students. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Shriner, J. G., Kim, D., Thurlow, M. L., & Ysseldyke, J. E. (1992a) Experts' opinions on national math standards for students with disabilities. Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Shriner, J. G., Kim, D., Thurlow, M. L., & Ysseldyke, J. E. (1992b) Experts' opinions on the appropriateness and feasibility of national math standards for students with disabilities. Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Shavelson, R., McDonnell, L., & Oakes, J. (1989). Indicators for monitoring mathematics and science education. Santa Monica, CA: RAND.

Southeastern Community College v. Davis, 442 U. S. 397, 413 (1979).

Spady, W. G. (1992). It's time to take a close look at outcome-based education. Communique, 20 (6), 16-18.

White House, 1990. National education goals. Washington, DC: Office of the Press Secretary.

Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). Testing handicapped people. Boston: Allyn & Bacon.

Ysseldyke, J. E., Thurlow, M. L., Bruininks, R. H., Gilman, C. J., Deno, S. L., McGrew, K. S., & Shriner, J. G. (1992). An evolving conceptual model of educational outcomes for children and youth with disabilities (Working Paper 2). Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Ysseldyke, J. E., Thurlow, M. L., Bruininks, R. H., Deno, S. L., McGrew, K. S., & Shriner, J. G. (1991). A conceptual model of educational outcomes for children and youth with disabilities (Working Paper 1). Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Ysseldyke, J. E., Thurlow, M. L., & Shriner, J. G. (1992). Outcomes are for special educators too. Teaching Exceptional Children, 25 (1), 36-50.

Table 1

Technical Issues Identified by State Assessment Personnel

Issue Category	Description of Issues
Cross-sectional vs Longitudinal Design	Is one method preferred over the other for assessments including students with disabilities?
Data Aggregation	Should data collected as part of the federal special education child count be included with other data on students with disabilities? Should these data be integrated with the data of other students? How should data be aggregated when they are from differing data collection methods, differing measures, or differing metrics?
Equity	Is it fair to make special changes in assessment for one group of students, specifically those with disabilities? Is it fair not to make special changes in assessment for them?
Feasibility	Is it feasible to develop the new or modified scales that would be needed to include all students in assessments? Could specially selected subsamples of students with disabilities be included and their scores put on the same scale as the rest of the sample?
Instrument Adaptation	What accommodations can be made that will allow students with disabilities to participate in assessments? To what extent will it be permissible to allow teachers freedom to make judgments about the need to adapt or modify instruments to allow inclusion?
Out-of-Level Testing	When and how should out-of-level testing be used (e.g., giving a 9th grade student a 3rd grade test)? If used, how are these data included in aggregations?
Range of Items	How do we develop a wide enough range of items to include all students in an assessment?
Reliability	How can reliability of measures be assured -- how do we maintain consistency in instrument modifications from one year to the next?
Sampling Methodology	How do selected sampling techniques affect the inclusions of students with disabilities? Does over-sampling always have to be done to make up for the low rates of students with disabilities in the general population?

Table 1--continued

Technical Issues Identified by State Assessment Personnel

Issue Category	Description of Issues
Terminology	Jargon and definitional confusion impede implementation of a system of outcome indicators and aggregation of data. Terminology is a problem specific to the assessment of students with disabilities because there is inconsistent definition of disabilities across states.
Test Standardization	Standardization samples typically exclude students with disabilities, contributing to difficulties in later including students with disabilities in assessment practices.
Validity	Is there a way to assure validity of measures for which modifications have been made?

Note: Secondary technical issues are listed alphabetically here. Order does not reflect in any way, the number of individuals raising the issue.